Honda Research Institute EU
Honda Research Institute Europe GmbH
https://www.honda-ri.de/

Analysis of a Speech-Based Intersection Assistant in Real Urban Traffic

Dennis Orth, Nico Andreas Steinhardt, Bram Bolder, Mark Dunn, Dorothea Kolossa, Martin Ernst Heckmann

2018

Preprint:

This is an accepted article published in The 21st IEEE International Conference on Intelligent Transportation Systems. The final authenticated version is available online at: https://doi.org/[DOI not available]

Analysis of a Speech-Based Intersection Assistant in Real Urban Traffic

Dennis Orth[1,2], Nico Steinhardt[1], Bram Bolder[1], Mark Dunn[1], Dorothea Kolossa[2], Martin Heckmann[1]

[1]Honda Research Institute Europe, Offenbach/Main, Germany
[2]Institute of Communication Acoustics, Ruhr-Universität Bochum, Bochum, Germany

Email: {martin.heckmann, nico.steinhardt, bram.bolder, mark.dunn}@honda-ri.de,
{dennis.orth, dorothea.kolossa}@rub.de

Preprint version. Paper published in: The 21st IEEE International Conference on Intelligent Transportation Systems. © 2018 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

Analysis of a Speech-Based Intersection Assistant in Real Urban Traffic

Dennis Orth*†, Nico Steinhardt†, Bram Bolder†, Mark Dunn†, Dorothea Kolossa*, and Martin Heckmann†
*Ruhr University Bochum, Faculty of Electrical Engineering and Information Technology,
Institute of Communication Acoustics, Germany, Bochum
Email: {dennis.orth, dorothea.kolossa}@rub.de
†Honda Research Institute Europe GmbH, Germany, Offenbach am Main
Email: {nico.steinhardt, bram.bolder, mark.dunn, martin.heckmann}@honda-ri.de

***Abstract*—Previously, we have presented a speech-based intersection assistant prototype. The system is activated on-demand by the driver and gives afterwards, via speech, information on suitable gaps between the traffic vehicles approaching from the right. It is comparable to a front seat passenger which helps in the maneuver decision for an intended turn left. This system has assumed a more or less constant flow of the traffic. To also handle situations of more dynamic urban traffic, including vehicles that may be slowing down or stopping, we have now extended our previous approach by a dynamic vehicle model. This model predicts the future traffic vehicle state based on second-order vehicle dynamics. We perform an in depth analysis of our system on a set of recordings under various traffic conditions. In this analysis we compare in particular the previous and the novel vehicle model. Both approaches lead to a correct recommendation in approximately 90% of the cases. Unexpectedly, the dynamic model does not lead to significant improvements in the system behavior, despite its increased accuracy.**

I. Introduction

As stated in [1], [2], intersections are among the most confusing and complex places in traffic, which often leads to accidents. Therefore, with increasing technical possibilities in the automotive industry, intersection assistance systems are investigated to mitigate or prevent crashes at intersections [3]. These assistance systems often cover situations with oncoming traffic [4], [5], but are now being extended to also handle multi-directional collisions [6], [7]. Currently, those systems mainly use LIDAR and RADAR as environment perception technology. Thus the assistance systems are limited to emergency situations only. It is expected that this disadvantage will be mitigated in the future by applying vehicle-to-infrastructure and vehicle-to-vehicle communication approaches, which are under current research [8], [9], [10], [11]. Another reason for the limited operational range of the previous intersection assistants (mentioned above) is the rather low capability of interpreting the surrounding environment. Thus, one can find plenty of research with regard to traffic participants' maneuver recognition [12] and intention prediction [13], [14], [15], [16], [17].

However, the previous mentioned assistance systems still occasionally provide warnings when there is no danger. Additionally, often the drivers have already perceived the danger and have adapted their maneuver planning accordingly. In both cases the warnings of the systems may be perceived as annoying and may lead to the driver ignoring them or turning off the system. To overcome these limitations, we recently proposed the "Assistance on demand" (AOD) concept [18]. Particularly challenging for drivers is turning left at an unsignalized intersection from a subordinate road into a superordinate road, especially when the traffic density is high [19]. In the AOD approach the idea of a cooperative front seat passenger is transferred to the interaction with the assistance system. The driver can activate it via speech and it will give feedback, also via speech, on the current traffic situation to support her in the maneuver decision. Using speech the system will inform the driver on possible gaps in the traffic to make the turn. In [20] a similar approach for the use case of "turning left at a rural road with oncoming traffic" was applied in a simulator with positive results for the system. In contrast, the AOD approach is applied to a more demanding use case where a collaborative sharing of the driving task is more beneficial. In addition, the AOD system is activated on demand therefore minimizes possible annoyance. In [21] and [22] systems are proposed which also assist in decision making for oncoming traffic and crossing traffic, respectively, by providing Head-up display (HUD) based support. Yet, in [18] we could demonstrate in a simulator study that a system based on the AOD approach was well accepted by the participants and preferred compared to driving without assistance or with a HUD based support. To further improve the utility of the system, we have investigated individual drivers' gap acceptance and developed methods to efficiently estimate personalized gap recommendations [23], [24]. We could then show that these personalized recommendations clearly improved the acceptance of the system compared to identical recommendations for all drivers. They also enhanced the monitoring of the traffic situation and further decreased the perceived workload [25].

In [26] we implemented the previously described speech-based on-demand intersection assistant in a prototype vehicle and tested it in real urban traffic to investigate which additional challenges arise there for the design of the system. Based on the evaluations, we were able to design a system which enables tailored assistance while avoiding annoyance already in a wide range of traffic situations. In this work we provide a detailed analysis of one of the core elements of our system,

namely the vehicle behavior estimation and prediction. It is the basis for the gap estimation and therefore determines what will be proposed to the driver. In [26] we assumed that the traffic vehicles will drive at constant speed. However, in our experiments we observed that vehicles frequently had to slow down and stop, e.g. due to pedestrians which were crossing the street. This behavior is not covered by the constant velocity model in [26] and leads to an underestimation or overestimation of the gaps. The latter is the more critical case, as it could confuse the driver and generate risky situations. To overcome these limitations, we have also included in our analysis a novel vehicle model based on second-order dynamics. In Sec. II we give a short overview of the system's components and how they work together to accomplish the desired assistance functionality. The dialog manager, which controls the interaction with the driver, is described in Sec. III. It is followed by the description of the two vehicle behavior estimation approaches (Sec. IV). After explaining our experimental setup in Sec. V, we will show and discuss the results of our analysis (Sec. VI, Sec. VII).

II. System Overview

Our intersection assistant in [26] uses LIDAR sensors to estimate the position and the velocity of all traffic participants and based on that calculates the gaps between the approaching vehicles. An additional vehicle tracking method was implemented to mitigate frequent occlusions of traffic vehicles. The ego vehicle state estimation module uses vehicle CAN data to detect the arrival and departure of the ego vehicle at the intersection, to only provide feedback via the text-to-speech module while the driver is standing at the intersection. The speech commands from the driver are acquired via a microphone array and forwarded to the automatic speech recognition (ASR). The latter listens for the wake-up word "Cora" and triggers the recognition after the detection. If the natural language understanding (NLU) component detects the intent "watch right" in the recognition result, the dialog manager is activated. The dialog manager then continuously receives the data from the gap estimation and ego vehicle state estimation module, analyzes it and informs the driver accordingly.

III. Dialog Manager

One of the key features of the AOD system is the situated dialog, i.e., a dialog which is targeted at and embedded in the physical environment [27]. Our scenario is characterized by a highly dynamic environment due to the relatively rapidly moving vehicles. This requires a predictive dialog planning where the future state of the environment has to be considered in relation to the time required for speech synthesis and the average time a driver needs for listening to and understanding the message. This prediction is limited mainly by the system's perception range and the accumulating uncertainties in the future environment state with an increasing prediction horizon. In the following we will detail the dialog management approach we devised to tackle the aforementioned problems.

A. General Dialog Logic

References of the system to traffic participants have to be unambiguous as they might otherwise confuse the driver. Due to sensory limitations only the traffic participants dynamics are known. With current sensors, additional features such as type (car, truck, pedestrian ...), size or color are not available, or can only be obtained at too late a point in time. Hence the system refers to all traffic participants as "vehicles." To achieve the required unambiguous reference, the system always makes reference to the traffic participant approaching from the right with the shortest time of arrival. We will refer to this vehicle as the *trigger vehicle* (Fig. 1). In its approach several conditions are tested for the trigger vehicle, which can trigger an announcement of the system. We refer to the vehicle following the trigger vehicle as the *target vehicle* as the announcement is in most cases targeted at this vehicle. The reference point for the calculation of the time of arrival is in the center of the ego vehicle. We call this time t_{trigger} for the trigger vehicle. Consequently, we denote the time gap between the trigger and the target vehicle as t_{gap} and measure it from the front of the trigger vehicle to the front of the target vehicle. Due to aforementioned sensor limitations, we currently don't take the length of the trigger vehicle into account to determine the actual gap between the vehicles.

We will now illustrate the dialog logic in more detail based on a few possible situations. Assuming the trigger vehicle is still sufficiently far away to make an announcement ($t_{\mathsf{trigger}} > T_{\mathsf{announce}}$) and the gap behind this vehicle t_{gap} is large enough for the driver to make the turn, the system should announce "gap after next vehicle." In this utterance "next vehicle" refers to the trigger vehicle which triggered the announcement and "gap" to the target vehicle which is the vehicle which determines if the driver can make the turn. The driver is able to make the turn if $t_{\mathsf{gap}} \geq T_{\mathsf{gap\,crit}}$, where $T_{\mathsf{gap\,crit}}$ is the so called critical gap, a gap just large enough for the driver. As already mentioned above, the system needs time to output the utterance and the driver needs time to understand it, verify the traffic situation and take a decision. We accumulate this time in T_{announce}. Hence, the system should start with outputting "gap after next vehicle" at the latest when $t_{\mathsf{trigger}} = T_{\mathsf{announce}}$. Due to the limited and situation dependent perception horizon T_{predict}, it is advisable to make the announcement of the fitting gap as late as possible, i.e., when $t_{\mathsf{trigger}} = T_{\mathsf{announce}}$. This maximizes the likelihood that

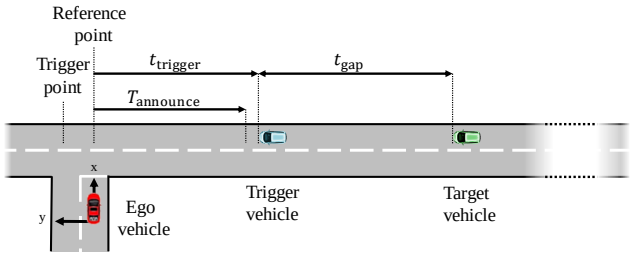

Fig. 1. Top view of the T-intersection, showing the ego vehicle (red) waiting for the vehicles from the right to pass the intersection.

all vehicles relevant for the planned announcement are already detected. In case $t_{\mathsf{gap}} < T_{\mathsf{gap\,crit}}$ the system will announce "vehicle from the right." To avoid confusions, this announcement is only made after the trigger vehicle has passed the intersection. This condition is fulfilled when the front of the vehicle passes the trigger point, a point which is $l_{\mathsf{vehicle\,default}}$ to the left of the reference point. The value $l_{\mathsf{vehicle\,default}}$ represents a default vehicle length, which we use as the real vehicle length is difficult to obtain reliably from the LIDAR. The next announcement will only be possible after the trigger vehicle has passed the trigger point. After this it is removed from the environment representation and the vehicle with the then shortest time of arrival is chosen as the new trigger vehicle.

IV. Traffic Vehicle Behavior Estimation

In the following we will describe the two models for the vehicle behavior estimation and prediction. At first we will introduce the second-order vehicle dynamics model which we use to describe and predict the vehicle's behavior. After that we will show how this vehicle dynamics are used to extend the behavior prediction compared to the constant velocity model.

A. Vehicle Dynamics

All position and velocity data which is delivered by the sensors are given in the coordinate system depicted in Fig. 1. Hence, the ego vehicle denotes the origin. Since we observed in our recordings that the traffic vehicles' velocity data is very noisy with regard to the x-component, we applied the following transformation in order to reduce the impact of that noise in the subsequent calculations: As a first step, the coordinates of the velocity vector $\mathbf{v}(k)$ are rotated into the body coordinates of the traffic vehicle $\mathbf{v}_b(k)$, with k as the current time stamp. It is expected that the x-axis of this rotated vector is in parallel to the crossing superordinate road. Thus the x-coordinate describes now the velocity in forward direction of the traffic vehicle. Furthermore, as that noise in the x-component can lead to negative velocities, we limited the minimum value of the x-coordinate to zero. To obtain the current *body* acceleration we calculate the difference quotient:

$$\mathbf{a}'_b(k) = \frac{\mathbf{v}_b(k) - \mathbf{v}_b(k-1)}{T_s}\,, \tag{1}$$

where T_s is our sampling interval. To circumvent erroneous acceleration estimations, we perform the step in Eq. 1 only when the vehicle is observable, thus the LIDAR sensors deliver real velocity estimates. However, since the latter contains also noisy values, the acceleration could still take unrealistic values which would finally lead to erroneous predictions of the gaps and arrival times. To prevent that, we also applied a limitation of the body acceleration x-component with:

$$a''_{b,x}(k) = \begin{cases} a_{\mathsf{max}} & \text{if } a_{b,x}(k) > a_{\mathsf{max}} \\ a_{\mathsf{min}} & \text{if } a_{b,x}(k) < a_{\mathsf{min}} \\ a'_{b,x}(k) & \text{otherwise} \end{cases} \tag{2}$$

In order to reduce remaining noise we finally smooth this acceleration with a low pass window-based (Hamming) FIR filter of length l_{FIR}. In the following we will denote with $a_{b,x}(k)$ the filtered acceleration $a''_{b,x}(k)$.

B. Behavior Prediction

One important value for our analysis is the predicted time t_{trigger}. This is the time the vehicle needs to arrive at the reference point $\mathsf{p}_{\mathsf{y,ref}}$. This arrival time is also the basis for calculating the gap the system proposes. Another important value is the predicted distance d_{predict}. It is defined as the distance that the vehicle will cover in T_{announce}. It will serve as a quantitative measure in the analysis to evaluate the precision of both models. In the following we will first describe the constant velocity model and show afterwards the extension to the dynamic model. For all calculations involving the body velocity or body acceleration, we use only the x-component of the vectors, thus $v_{b,x}$ and $a_{b,x}$.

1) Constant Velocity Model: This model uses the current velocity and the current position $\mathsf{p}_y(k)$ of the traffic vehicle to estimate d_{predict} and t_{trigger} as follows:

$$d_{\mathsf{predict}} = v_{b,x}(k) \cdot T_{\mathsf{announce}}, \quad t_{\mathsf{trigger}} = \frac{\mathsf{p}_y(k) - \mathsf{p}_{y,\mathsf{ref}}}{v_{b,x}(k)} \tag{3}$$

If the velocity $v_{b,x}$ falls below v_{stand}, t_{trigger} is set to ∞ and the predicted distance d_{predict}, which the vehicle will cover in T_{announce} seconds, is set to zero.

2) Dynamic Model: In the calculation of the dynamic model we have to differentiate between 4 cases: Whether the vehicle is decelerating, driving at constant speed, accelerating or standing. To prevent that small noisy values of the calculated acceleration induce faulty predictions, we only trigger the dynamic model when $-0.1 < a_{b,x}(k) < 0.1$, otherwise we use the constant velocity model to calculate the desired variables as in Eq. 3. This also includes the case of standing where we apply the same procedure as for the constant velocity model.

To compute the distance s that will be covered in time t, second-order motion dynamics are used via:

$$s(t) = v_{b,x}(k) \cdot t + 0.5 \cdot a_{b,x}(k) \cdot t^2\,. \tag{4}$$

In the deceleration case, at first the time the vehicle will need to stop t_{stop} and the corresponding distance d_{stop} which it will cover within this time, are calculated as follows:

$$t_{\mathsf{stop}} = -\frac{v_{b,x}(k)}{a_{b,x}(k)}, \quad d_{\mathsf{stop}} = s(t_{\mathsf{stop}})\,. \tag{5}$$

With that, we can compute the predicted distance within T_{announce} seconds.

$$d_{\mathsf{predict}} = \begin{cases} s(T_{\mathsf{announce}}) & \text{if } t_{\mathsf{stop}} \geq T_{\mathsf{announce}} \\ d_{\mathsf{stop}} & \text{otherwise}\,, \end{cases} \tag{6}$$

If the predicted stopping position is beyond the reference point we have to solve the equation

$$0 = 0.5 \cdot t^2 \cdot a_{b,x}(k) + v_{b,x}(k) \cdot t + (\mathsf{p}_y(k) - \mathsf{p}_{y,\mathsf{ref}}) \tag{7}$$

to obtain the time t_{trigger}. Otherwise it will stop before reaching the reference point, so t_{trigger} is set to ∞.

For the acceleration case we check if the vehicle has reached the maximum assumed velocity v_{max}. If not, it is assumed that the vehicle will continue accelerating until the maximum velocity is reached and that it will continue at a constant speed v_{max} afterwards. Therefore we calculate the time needed to reach v_{max} and the corresponding distance covered in that time:

$$t_{<\mathsf{v}_{\mathsf{max}}} = \frac{v_{\mathsf{max}} - v_{b,x}}{a_{b,x}}, \quad d_{<\mathsf{v}_{\mathsf{max}}} = s(t_{<\mathsf{v}_{\mathsf{max}}}) \,. \tag{8}$$

After that $t_{\mathsf{remain}} = T_{\mathsf{announce}} - t_{<\mathsf{v}_{\mathsf{max}}}$ is calculated and depending on that result it controls the following calculation:

$$d_{\mathsf{predict}} = \begin{cases} d_{<\mathsf{v}_{\mathsf{max}}} + v_{\mathsf{max}} \cdot t_{\mathsf{remain}} & \text{if } t_{\mathsf{remain}} \geq 0 \\ s(T_{\mathsf{announce}}) & \text{otherwise} \,, \end{cases} \tag{9}$$

If $\mathsf{p}_{y,v_{\mathsf{max}}}(k) = \mathsf{p}_y(k) + d_{<\mathsf{v}_{\mathsf{max}}}$ is beyond the reference point we calculate t_{trigger} again via Eq. 7, otherwise:

$$t_{\mathsf{trigger}} = t_{<\mathsf{v}_{\mathsf{max}}} + \frac{\mathsf{p}_{\mathsf{ref}} - \mathsf{p}_{v_{\mathsf{max}}}}{v_{\mathsf{max}}} \,. \tag{10}$$

If the vehicle has already reached v_{max}, we assume that, despite measuring $a_{b,x}(k) > 0$, the vehicle will not further accelerate, but continue with the current speed. In this case t_{trigger} and d_{predict} are calculated again with Eq. 3.

C. Gap Estimation

Relevant for our system are only those vehicles which are approaching from the right side, since the driver is still monitoring the left side. To identify these vehicles reliably we apply two criteria. First, in addition to being on the right side of the ego vehicle they have to have a heading angle in the range of $[180°, 0°]$. This angular range has to be wide enough to allow for slightly tilted positioning of the ego vehicle at the intersection, non-orthogonal T-intersections and moderate errors in the tracking of the surrounding vehicles. Second, they have to arrive towards the ego vehicle in the angular ranges $[0°, -130°]$ or $[0°, 130°]$. The latter filters vehicles which approach from behind (see Fig. 1). For the remaining vehicles the values t_{trigger} and d_{predict} are calculated as described in Sec. IV-B. However, the list of current vehicles is then sorted by the current positions and not by the arrival times t_{trigger}. Based on this list, the time gaps between the vehicles are computed.

V. Experimental Setup

To analyze our system, we have recorded a small dataset in real urban traffic at an intersection in Offenbach (Germany). A modified 2012 model-year Honda CR-V was used as the prototype vehicle. In addition to the standard equipment, it features 6 Ibeo Automotive Systems LUX LIDAR sensors, which have an overall coverage of 360°. The LIDAR raw data is processed in a dedicated Ibeo sensor fusion and object detection & tracking unit. The LIDAR has a sampling rate of $f_{\mathsf{LIDAR}} = 25\,\mathrm{Hz}$ which drives the vehicle tracking and gap estimation.

The prototype vehicle was standing still at the intersection, as depicted in Fig. 1. All recordings were then passed through the vehicle tracking and gap estimation component of the scene understanding, which are described in Sec. IV. Since the system is only proposing gaps for the right side, only those vehicles approaching from the right were taken into account. In addition, all vehicles that were present for less than T_{announce} (2.5 seconds) were also not taken into account since our smallest prediction horizon to be evaluated is T_{announce}. After all, 133 vehicles remained which could be used for the analysis. To generate a gold standard for the analysis, we applied the following approach: We used the position which was delivered from the scene understanding component and linearly interpolated the position signal for those segments where the vehicle was not observable, provided that this part was surrounded by time segments, during which the vehicle was observable. For all segments where the vehicle was observable, the position corresponded to the raw position delivered by the IBEO sensory system. Non-causal smoothing of the position estimate was performed using a Hamming window with a size of 51 samples (2.04 s). The first 25 samples in the beginning of the detection of the vehicle, i.e. the setting time of the smoothing filter, were not smoothed.

As described in Sec. V, we denote the time the driver needs for hearing and understanding the messages by T_{announce}. Therefore we evaluate the constant velocity and the dynamic vehicle model with regard to the *predicted position* of the traffic vehicles which they will have after T_{announce} seconds.

A reasonable approach is to check whether the announcement that was given when the trigger vehicle was T_{announce} seconds away, is still valid when this vehicle reaches the reference point in front of the ego vehicle. For this analysis of the *gap prediction*, more vehicles had to be filtered out of the pool: We only considered trigger vehicles that did not pass the intersection at the beginning of the recording, whose t_{trigger} fell below T_{announce} at some point and which finally reached the reference point ($t_{\mathsf{trigger}} \leq 0$). In addition, if those trigger vehicles had a following target vehicle, the latter also had to pass the intersection. Otherwise the announcement could not be evaluated. With these criteria, 115 vehicles were extracted for the gap prediction analysis. $T_{\mathsf{gap\,crit}}$ was set to $6.0\,\mathrm{s}$. The procedure for checking whether the announcements are still valid was to extract the order in which the vehicles drove through the reference point, calculate the gap sizes via the timestamps and compare the actual gaps with the output of the algorithms.

Three types of announcements could occur in the context of the previously described dialog system: *Announce No Vehicle* means that there is no following vehicle at all. With *Announce Gap* we denote the situation $t_{\mathsf{gap}} \geq T_{\mathsf{gap\,crit}}$ and with *Announce Nothing* $t_{\mathsf{gap}} < T_{\mathsf{gap\,crit}}$. In addition to the previous check, it is necessary to ascertain that the ID of the actual target vehicle is the same as the one, for which the announcement was made. These IDs can differ due to sensor issues, which may lead to new vehicles suddenly occurring between the reference point and the current target vehicle, or to target vehicles that suddenly vanish before they pass the intersection.

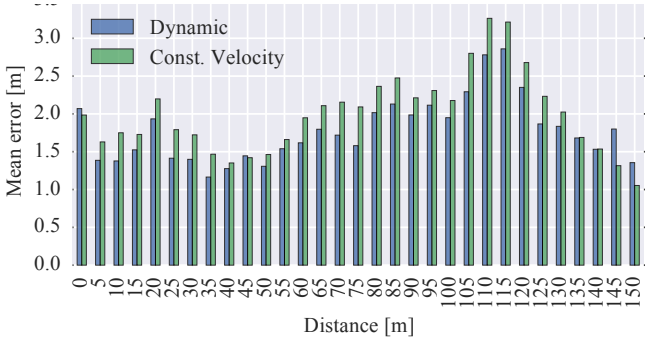

Fig. 2. The x-axis denotes the location of the traffic vehicle, when the prediction was made. The corresponding bars show the mean absolute error after T_{announce} seconds.

VI. Analysis

In this section we will present the results of the experiment described in Sec. V.

A. Position Prediction

The mean absolute prediction error vs. the distance of the traffic vehicle is shown in Fig. 2. Note that the error plotted on the y-axis denotes the error the system produces T_{announce} seconds in the future, and not at the location given by the x-axis. Hence, the actual position where the error arises will be closer to the reference point. The maximum mean error of $3.33\,\text{m}$ appears at around $110\,\text{m}$ for the constant velocity model. For the dynamic model, the maximum mean error of $2.95\,\text{m}$ occurs for distance of $115\,\text{m}$. Assuming a velocity of $50\,\text{km}\,\text{h}^{-1}$, the values correspond to $0.239\,\text{s}$ and $0.212\,\text{s}$. With that velocity, t_{trigger} will fall below T_{announce} when the trigger vehicle is approximately $35\,\text{m}$ away. For this distance we obtain mean errors $1.2\,\text{m}$ ($0.086\,\text{s}$, Std.: $1.2\,\text{m}$) for the dynamic model and $1.5\,\text{m}$ ($0.072\,\text{s}$; Std.: $1.3\,\text{m}$) for the constant velocity model. If we take the mean over all locations we obtain $1.72\,\text{m}$ (Std.: $1.86\,\text{m}$) for the dynamic and $1.97\,\text{m}$ (Std.: $2.04\,\text{m}$) for the constant velocity model.

B. Gap Prediction

We consider an announcement, which was made by the system, as valid if it matches the actual traffic situation or at least would lead to the same decision of the driver. E. g. if the system announced no following vehicle, yet suddenly a vehicle arrives in the sensor range, we do not count it as an error if the gap to this vehicle is large enough for the driver to make the turn. The constant velocity model led to valid announcements in 103 cases (89.6%) while the dynamic model led to valid announcements in 104 cases (90.4%). Fig. 4 and Fig. 5 show the number of valid and invalid announcements. At every bar on the y-axis, the first condition stands for the predicted traffic situation while the second (after the semicolon) stands for the one that actually occurred.

1) Valid Announcements: As depicted in Fig. 3, 29% of the gaps in the recordings were greater or equal to the critical gap ($t_{\text{gap}} \geq T_{\text{gapcrit}}$). Thus a lot of *Announce Nothing* cases appear in our recorded data. As one can see, most of these cases were

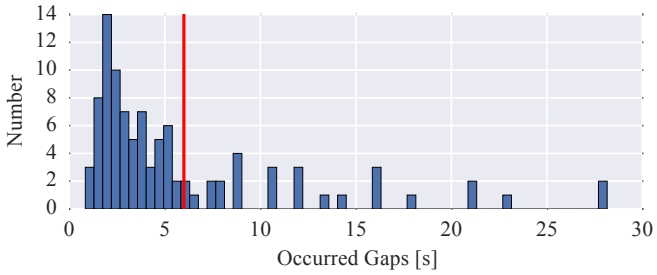

Fig. 3. This figure shows a histogram of all actual occurred gaps in our used recordings. The vertical red line depicts the critical gap which had a value of $6.0\,\text{s}$.

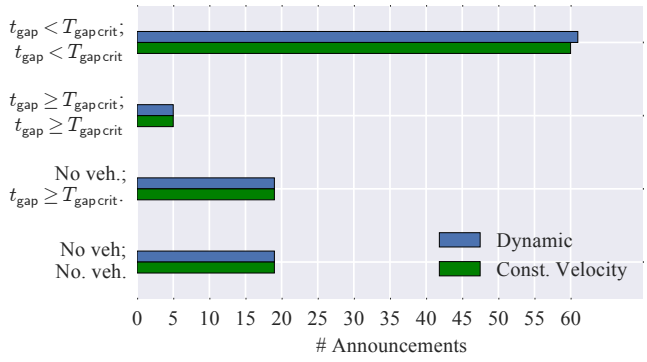

Fig. 4. Analysis of the given announcements that were valid.

correctly predicted by the system (Fig. 4, topmost bar). The constant velocity model predicted in this case once more a suitable gap that was actually not suitable, compared to the dynamic model. There are 24 occurrences of suitable gaps, which were always correctly classified. 5 of those times, they led to the decision *Announce Gap* and 19 times as *Announce No Vehicle*. We count the latter as valid since, at the time of the announcement, the vehicles were outside our sensor range and it should lead to the same, correct, decision of the driver, who should proceed to take the turn. In 19 times it was predicted that there is no following target vehicle

2) Invalid Announcements: The most critical situation, namely, the announcement of a suitable gap that was actually too small, appeared for both models, but one time more often for the constant velocity model (Fig. 5). For the remaining incorrect announcements, both models behaved identically. In one case, the ID of the actual target vehicle was different from the ID of the announcement. This can also be classified as a problem of the sensory system. Three times the system actually failed to propose a suitable gap.

VII. Discussion

As described in Sec. VI, both models already delivered precise prediction results for the position of the traffic vehicles in T_{announce} seconds. Especially for the area where the announcement is most likely made, the mean absolute errors were small enough to obtain a well timed announcement. For example, if the announcement is triggered too late, the driver probably will not have enough time to take a decision and perform his action. However, the dynamic model delivered slightly better results.

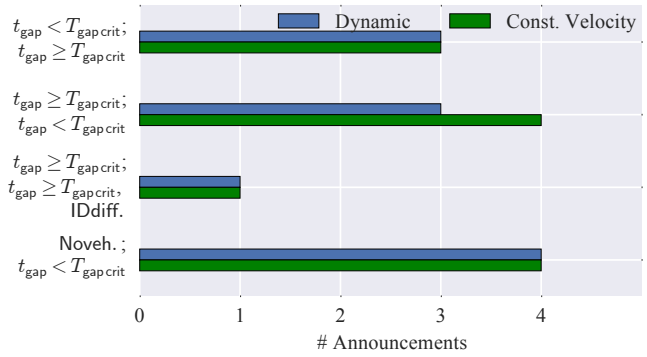

Fig. 5. Analysis of the given announcements that were invalid.

Fig. 4 and Fig. 5 show that both models already lead to valid announcements for approximately 90% of all cases in our test dataset. Unexpectedly, the dynamic model did not yield significant improvements in the gap prediction. As shown in Fig. 3, a large percentage of the actual gaps lay below $5.0\,\mathrm{s}$. The predicted gaps must vary significantly to have an effect on the result of the binary decision. Due to the large prediction horizon in combination with the highly dynamic traffic scenario this only occurs rarely. Only when the traffic vehicles accelerate or decelerate constantly for a longer time period, a higher benefit can be gathered. A clear constant behavior in the acceleration is not observed very often. It appears that typically, the driver will rather decelerate or accelerate for a short time and will afterwards drive at constant speed. Another reason for the rather low benefit from the dynamic model is the noisy velocity estimation and the resultant, even noisier acceleration estimation. The filtering, that this necessitated, induced a certain delay. This might be tackled with an improved sensor system, which delivers more precise velocity data.

VIII. Conclusion

In this work we have extended our recently proposed speech-based intersection assistant in regard to its traffic vehicle behavior estimation and prediction component by introducing a dynamic vehicle model. This system involves second-order vehicle dynamics to describe and predict the vehicle behavior. Additionally, we have performed an analysis of our system using the dynamic and our previous constant velocity vehicle model. As test data set we used real traffic data which was recorded at an urban intersection with our LIDAR equipped prototype car. The results have shown that using both models led to a correct behavior of the system in approximately 90% of the cases. Using the dynamic model yields no significantly better performance compared to the constant velocity model.

References

[1] Directorate General for Transport, "Traffic safety basic facts 2016: Junctions," European Commission, Tech. Rep., Jun. 2016.

[2] W. G. Najm, B. Sen, J. D. Smith, and B. N. Campbell, "Analysis of Light Vehicle Crashes and Pre-Crash Scenarios Based on the 2000 General Estimates System," Tech. Rep., 2003.

[3] J. M. Scanlon, R. Sherony, and H. C. Gabler, "Preliminary potential crash prevention estimates for an intersection advanced driver assistance system in straight crossing path crashes," in *2016 IEEE Intell. Vehicles Symposium*, June 2016, pp. 1135–1140.

[4] Volvo Cars. [Online]. Available: www.volvocars.com/us/about/our-innovations/intellisafe

[5] Audi AG. [Online]. Available: https://audi-dialoge.de/en/Kreuzungsassistent

[6] Mercedes-Benz. [Online]. Available: www.mercedes-benz.com/en/mercedes-benz/innovation/on-the-radar-screen-recognising-risks-automatically/

[7] M. Mages, F. Klanner, and A. Stoff, *Intersection Assistance*. Cham: Springer International Publishing, 2016, pp. 1259–1286.

[8] G. Karagiannis, O. Altintas, E. Ekici, G. Heijenk, B. Jarupan, K. Lin, and T. Weil, "Vehicular networking: A survey and tutorial on requirements, architectures, challenges, standards and solutions," *IEEE Communications Surveys Tutorials*, no. 4, pp. 584–616, 2011.

[9] V. A. Butakov and P. Ioannou, "Personalized Driver Assistance for Signalized Intersections Using V2I Communication," *IEEE Trans. on Intelligent Transp. Systems*, 2016.

[10] Mercedes. [Online]. Available: https://www.mercedes-benz.com/en/mercedes-benz/innovation/car-to-x-communication

[11] BMW. [Online]. Available: www.press.bmwgroup.com/global/article/detail/T0274952EN/bmw-motorrad-presents-the-r-1200-rs-connectedride

[12] Q. Tran and J. Firl, "Online maneuver recognition and multimodal trajectory prediction for intersection assistance using non-parametric regression," in *2014 IEEE Intell. Vehicles Symposium*, June 2014.

[13] C. Rodemerk, H. Winner, and R. Kastner, "Predicting the driver's turn intentions at urban intersections using context-based indicators," in *2015 IEEE Intell. Vehicles Symposium*, June 2015, pp. 964–969.

[14] A. Zyner, S. Worrall, and E. Nebot, "A recurrent neural network solution for predicting driver intention at unsignalized intersections," *IEEE Robotics and Automation Letters*, July 2018.

[15] M. Liebner, M. Baumann, F. Klanner, and C. Stiller, "Driver intent inference at urban intersections using the intelligent driver model," in *IEEE Intell. Vehicles Symposium*, 2012.

[16] A. Zyner, S. Worrall, J. Ward, and E. Nebot, "Long short term memory for driver intent prediction," in *2017 IEEE Intell. Vehicles Symposium*.

[17] V. Losing, B. Hammer, and H. Wersing, "Personalized maneuver prediction at intersections," in *IEEE 20th Int. Conf. on Intell. Transp. Syst. (ITSC)*, Yokohama, Japan, oct 2017.

[18] N. Schoemig, M. Heckmann, H. Wersing, C. Maag, and A. Neukum, ""Please watch right" - Evaluation of a speech-based on-demand assistance system for urban intersections," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 54, pp. 196 – 210, 2018.

[19] A. Stinchcombe and S. Gagnon, "Estimating Workload Demands of Turning Left at Intersections of Varying Complexity," *Proc. of the 5th Int. Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*, vol. 5, pp. 440–446, 2009.

[20] L. Biester, "Cooperative automation in automobiles," Ph.D. dissertation, Humboldt-Universität zu Berlin, 2009.

[21] C. Tran, K. Bark, and V. Ng-Thow-Hing, "A Left-Turn Driving Aid Using Projected Oncoming Vehicle Paths with Augmented Reality," in *AutomotiveUI 13*, 2013, pp. 300–307.

[22] B. Yang, R. Zheng, K. Shimono, T. Kaizuka, and K. Nakano, "Evaluation of the effects of in-vehicle traffic lights on driving performances for unsignalised intersections," *IET Intelligent Transport Systems*, 2017.

[23] D. Orth, D. Kolossa, M. Sarria Paja, K. Schaller, A. Pech, and M. Heckmann, "A maximum likelihood method for driver-specific critical-gap estimation," in *2017 IEEE Intell. Vehicles Symposium (IV)*.

[24] D. Orth, D. Kolossa, and M. Heckmann, "Predicting driver Left-Turn behavior from few training samples using a Maximum-A-Posteriori method," in *IEEE 20th Int. Conf. on Intell. Transp. Syst. (ITSC)*, 2017.

[25] D. Orth, N. Schömig, C. Mark, M. Jagiellowicz-Kaufmann, D. Kolossa, and M. Heckmann, "Benefits of personalization in the context of a speech-based left-turn assistant," in *AutomotiveUI '17*.

[26] D. Orth, B. Bolder, N. Steinhardt, M. Dunn, D. Kolossa, and M. Heckmann, "A speech-based on-demand intersection assistant prototype," in *2018 IEEE Intell. Vehicle Symposium*, June 2018.

[27] M. Heckmann, D. Orth, and D. Kolossa, ""Gap after the next two vehicles": A spatio-temporally situated dialog for a cooperative driving assistant," in *13. ITG Facht. Sprachkommunikation*. IEEE, October 2018.